Modeling Pathways of Cell Differentiation in Genetic Regulatory Networks With Random Boolean Networks
by Sheldon Dealy
Topics

- Introduction
- Background
- Biology
- Model
- Methods
- Results
- Discussion
- Summary
- Future Work
High Level View

- Can cells be modeled as non-linear dynamical systems -> cell differentiation?
- We chose ensembles of random Boolean networks to be the model.
- Ensembles of networks allowed us to find the typical behavior of the model.
- Results indicate phenomena which might occur in cell differentiation.
- These results are testable using gene arrays.
Introduction

- If cells can be modeled using non-linear dynamical systems, then cell types are attractors.
- If a cell type is an attractor, then the steps from one attractor to another is a pathway of cell differentiation.
- Hypothesis: Boolean networks share some key behaviors of cells in the process of differentiation.
Findings suggest Boolean networks are able to support some predictions about pathways of differentiation.

Predictions should be testable using gene array techniques.

Challenge: How to map the continuous nature of cell dynamics to the discrete steps of a synchronous Boolean network?
Early 1960's, Jacob and Monod proposed a “genetic circuit” to explain mechanics of cell differentiation.
If we restrict a gene's activity to being either on or off, then a cell can be modeled with a Boolean network.

Such a network having N genes has $2^N$ possible states.

The human genome has an estimated 25,000 genes.

A model network would then have $2^{25,000}$ possible states!
Biology

- Each gene is a segment of DNA.
- Interaction between genes induce them to express proteins which in turn can modify their own or another gene's behavior.
- These interactions form the connections of a genetic regulatory network.
- Genes in cells are in a continuous level of activation.
- Gene expression levels determines the cell's type.
Cell Differentiation

- Process where a cell of one type becomes a cell of another type.
- Cells differentiate as part of a normal process: embryonic stem cells can differentiate into one of 265 different cell types.
- Cells may also be induced to differentiate through exposure to certain chemicals.
- Gene arrays are used to take measurements of cell differentiation.
Random Boolean Networks (RBNs)

- RBNs are directed graphs of N nodes with K inputs to each node.
- Wiring of network is chosen at random.
- Boolean functions: Rules which allow for state transitions in the network.
- State of network is a snapshot of all node values.
- Each node evaluates to 1 or 0, on or off.
- Subsequent states generated synchronously by feeding the network state through a set of Boolean functions.
Model cont.

A simple K=2, N=3 Boolean network
A sequence of states which repeats itself is called a state-cycle or attractor.

All Boolean Networks have state cycles.

If cell types are attractors, then we can model them with RBNs.

As previously stated, getting from one attractor to another becomes a model for a pathway of differentiation.

In RBNs, a pathway of differentiation, or transient, is initiated by perturbing a state.
Testing Network Dynamics

• Needed to confirm the network dynamics for ordered, critical, and chaotic networks.
• The standard way to measure the complexity of a Boolean network was discovered by Derrida and Pommeau in 1986.
• Idea: compare the Hamming distance between pairs of states (\(dt\)) and their successor states (\(dt+1\)) in a random Boolean network.
Dynamics cont.

- If the Hamming distances are greater, then nearby states are diverging.
- If the Hamming distances are less, nearby states are converging.
- If the Hamming distances are about equal, then nearby states are neither converging nor diverging.
- Repeat for different pairs of states and plot the results.
Derrida plot
Experimental Methods

- Networks generated for various N,K values.
- For each state-cycle found, all single node perturbations generated to create transients.
- Statistics were gathered for:
  - Homeostatic versus non-homeostatic behavior.
  - Transient fusion, where pairs of transients join and flow together.
  - Hamming distances between states along the transient.
  - Number of times a node (gene) changes state over the length of a transient.
Homeostatic versus non-homeostatic behavior

- Homeostasis is a sign of stability.
- Networks with a single attractor excluded.
- For $N=10\text{--}40$, percentages of homeostatic transients decrease monotonically as $K$ increases for $K=1,2,3,4$.
- This is consistent with the idea that cell types are stable attractors.
Transient Length vs. Percentage of Transients

- Transient length is the number of state transitions required to reach an attractor.
- As K increases, the average transient length becomes longer.
- Homeostatic, but not non-homeostatic transients show a pronounced peak in transients of a single state transition.
  - We can predict the same for living cells.
Transient length vs homeostasis, K=2
Mapping the discrete to the continuous

- Idea: Track the incidence of gene change or “flips” over a transient.
- For K=2,3,4 the number of gene flips was found to increase monotonically.
Gene flips as a function of transient length
Transient fusion

- Fusion: Where transient paths merge and flow together.
- Is the process of convergence a smooth or sudden process?
- It was found that between pairs of transients which fuse, the Hamming distances converged smoothly with proximity to the point of fusion.
- Note: this cannot happen for K=N.
Convergence with transient fusion
Fusion Statistics

- It was found that as $K$ grows larger, the number of fused homeostatic transients grows smaller.
  - This indicates a higher convergence in state space for networks in the ordered and critical regime than for networks in the chaotic regime.
- It is hypothesized that the ratio of fused homeostatic to unfused homeostatic transients is a marker of the critical regime.
Fused and unfused transients
Hamming distances along the trajectory

- It was hypothesized that the hamming distance between successive states along the trajectory would monotonically decrease.
- This behavior was indeed found to exist for all K-values explored.
- Findings support a study done at Harvard Children's Hospital.
- Unfortunately the standard deviation is large, so the reduction in Hamming distance is difficult to notice.
Hamming distance along trajectory
Distribution of gene flips as a function of transient length

- The fraction of times along a transient that a gene changes state.
- K=1, 2 shows a wide distribution of the fraction of times genes change along transient.
- K=3, 4 shows that a large fraction of genes never change state.
- These features should be experimentally testable using a gene array.
Fraction of time genes flip along transient, K=1
Fraction of time genes flip along transient, $K=2$
Fraction of time genes flip along transient, $K=3$
Fraction of time genes flip along transient, $K=4$
The number of fusing pairs of transients discovered in RBNs suggests that fusing pathways of differentiation can be found.

A large percentage of genes in transients of chaotic networks never change state where the reverse is true in the ordered and critical regimes.

- If cells are stable attractors, we would expect to find a wide distribution in the percentage of time that genes change state.
Discussion
Discrete vs. Continuous

- The mapping of gene flips as a function of transient length suggests a linear relation between transient lengths in RBNs and the number of times a gene alters activity.
- If the mapping is valid, it should be possible using gene arrays to measure the number of gene variations over a series of closely timed intervals.
Summary

- This is the first examination of pathways of differentiation under the hypothesis that cell types correspond to attractors.
- It was not expected that RBNs would yield a comprehensive model of living cells.
- However, some behaviors in RBNs were found which should be measurable in living cells.
• Expected measurable features:
  – The amount of gene activity over time.
  – The ratio of fused homeostatic to unfused non-homeostatic transients.
  – The Hamming distance between successive states on a transient as it approaches a cell type.
  – The Hamming distance between pairs of transients as they approach a common cell type.
Future Directions

- **Medusa networks**
  - Small group of regulator genes, “head”.
  - Large group of regulated genes, “tail”.
  - Follows some behavior seen in living cells.

- **Asynchronous RBNs**
  - Updates genes one at a time, randomly.
  - Adds non-determinism to the model.
  - ARBNs may more closely approximate gene expression in genetic regulatory networks.
    - But nobody really knows
Acknowledgements

- I would like to thank Dr. Christopher Moore, Dr. Robert Veroff, and Dr. Stuart Kauffman for consenting to serve on my committee and graciously allowing me the use of their time.
- This research has been partially supported by grants from the National Science Foundation (PHY-0417660) and the National Institutes of Health (GM070600-01).